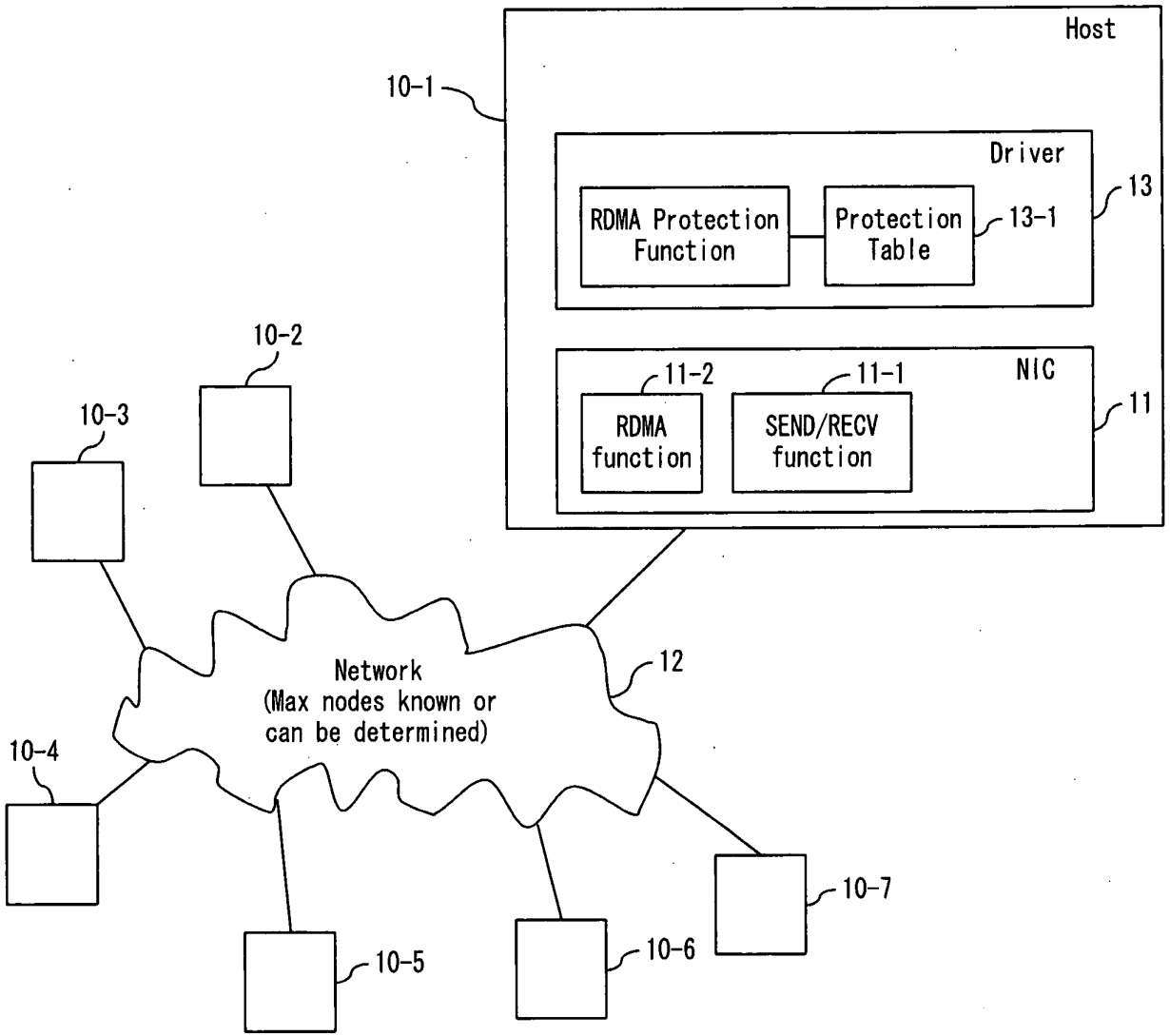
Host
10-1
Driver
13
RDMA Protection Function
Protection Table
13-1
10-2
11-2
11-1
NIC
10-3
RDMA function
SEND/RECV function
11
Network
(Max nodes known or can be determined)
12
10-4
10-7
10-5
10-6

FIG. 1

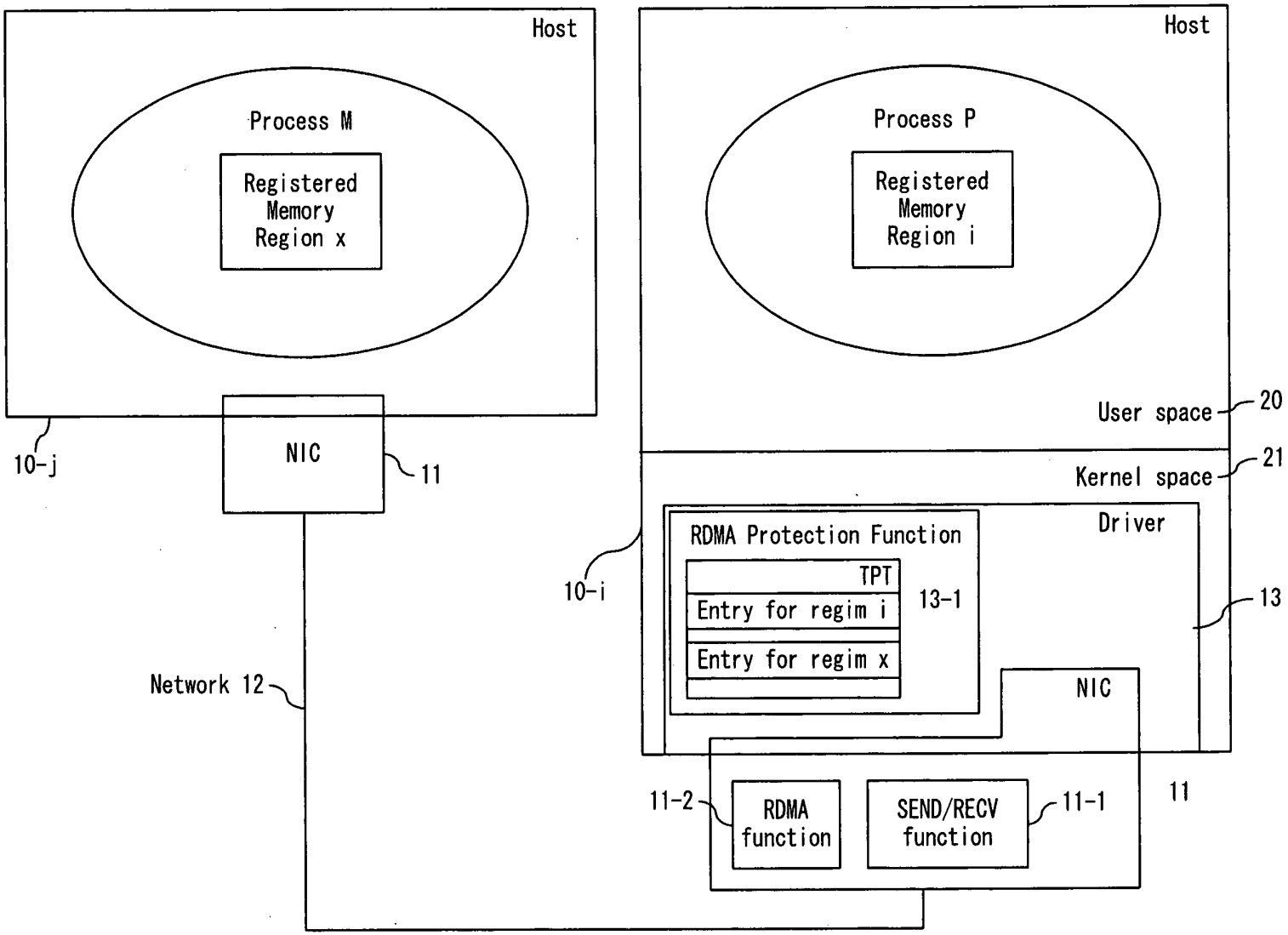
Host
Process M
Registered Memory Region x
10-j
NIC
11
Network 12
Host
Process P
Registered Memory Region i
User space
20
Kernel space
21
Driver
RDMA Protection Function
TPT
Entry for regim i
Entry for regim x
13-1
10-i
13
NIC
11-2
RDMA function
SEND/RECV function
11-1
11

FIG. 2

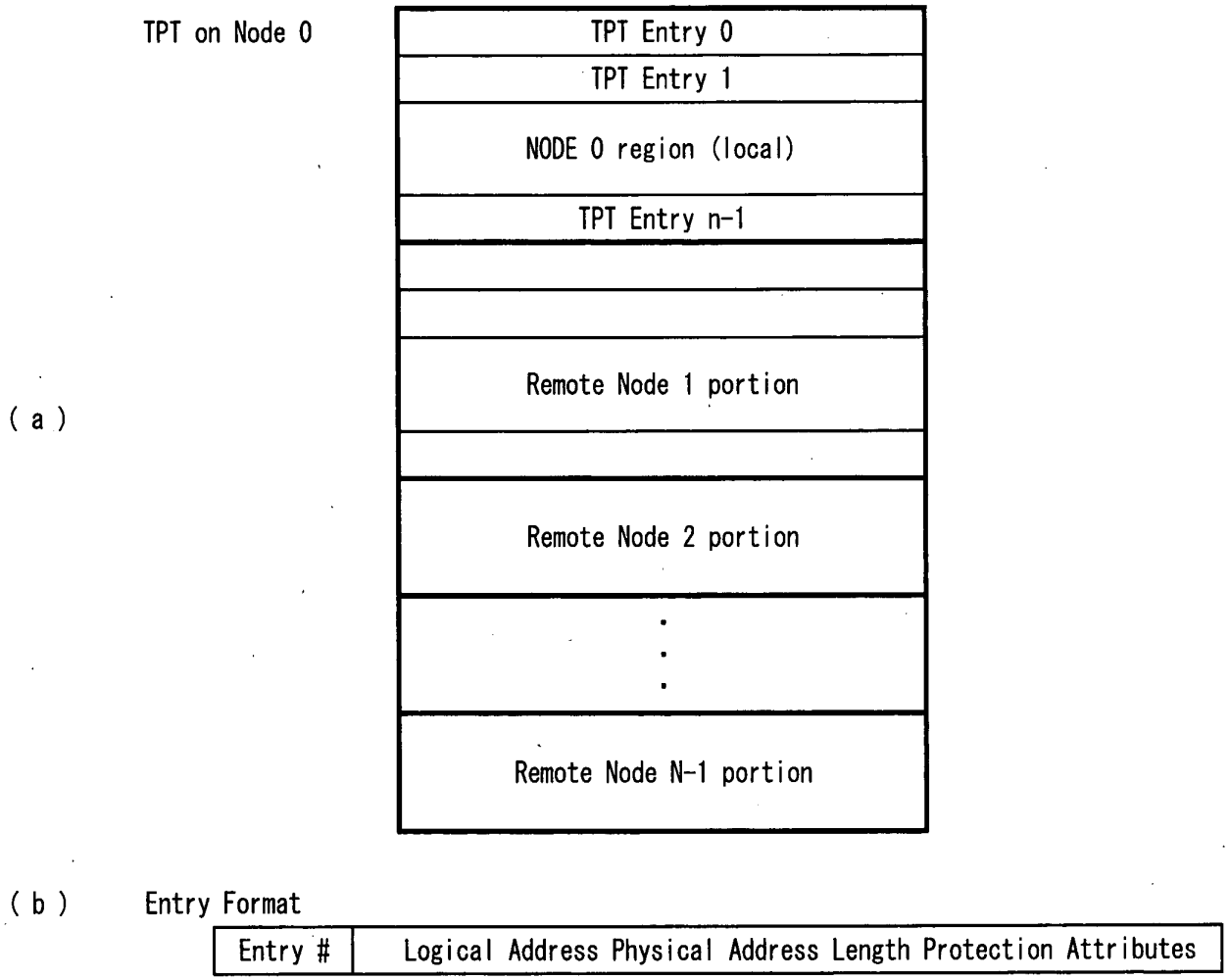
TPT on Node 0
TPT Entry 0
TPT Entry 1
NODE 0 region (local)
TPT Entry n-1
Remote Node 1 portion
Remote Node 2 portion
Remote Node N-1 portion
(a)
(b)
Entry Format
Entry #
Logical Address Physical Address Length Protection Attributes

FIG. 3

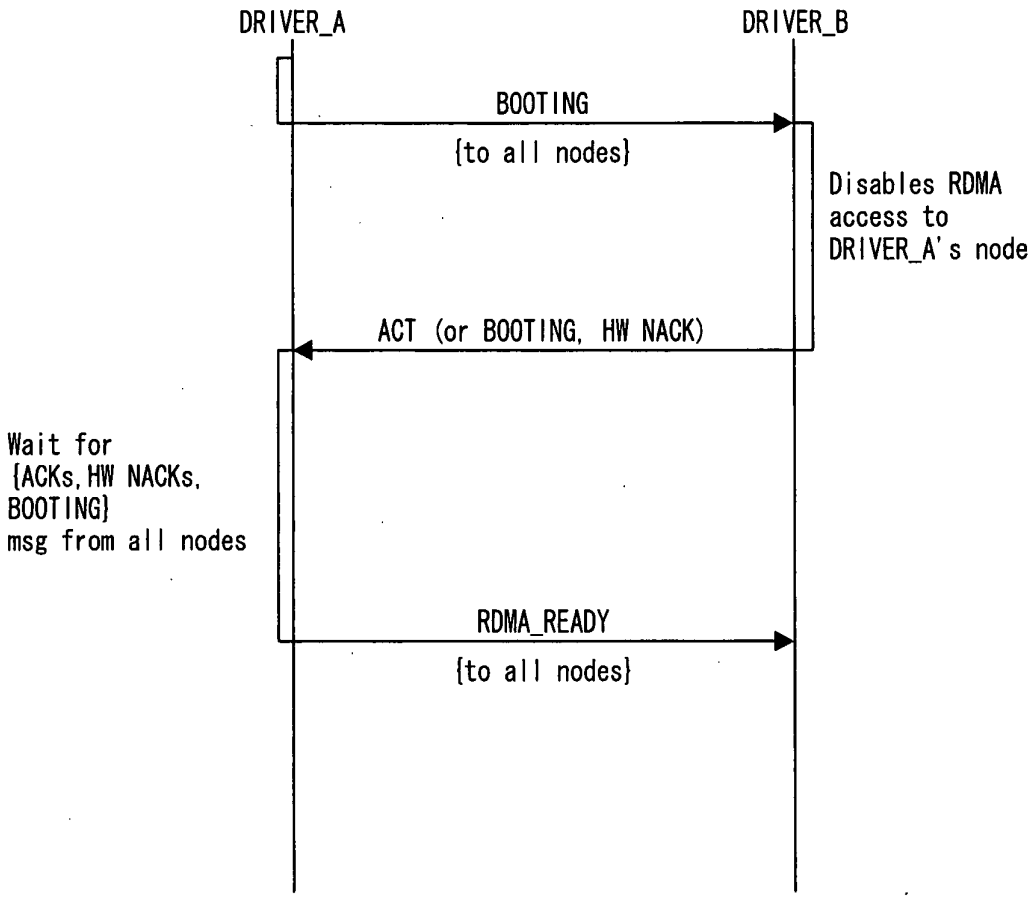
DRIVER_A
DRIVER_B
BOOTING
{to all nodes}
Disables RDMA
access to
DRIVER_A's node
ACT (or BOOTING, HW NACK)
Wait for
{ACKs, HW NACKs,
BOOTING}
msg from all nodes
RDMA_READY
{to all nodes}

FIG. 4

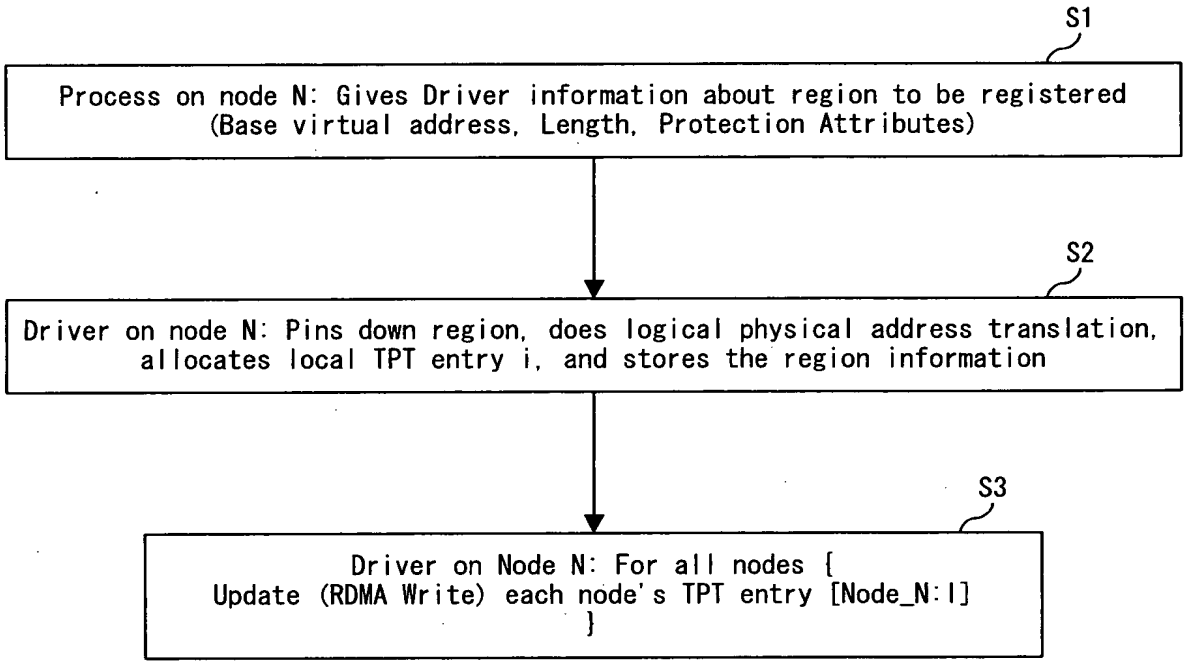
S1
Process on node N: Gives Driver information about region to be registered
(Base virtual address, Length, Protection Attributes)
S2
Driver on node N: Pins down region, does logical physical address translation,
allocates local TPT entry i, and stores the region information
S3
Driver on Node N: For all nodes {
Update (RDMA Write) each node's TPT entry [Node_N:I]
}

FIG. 5

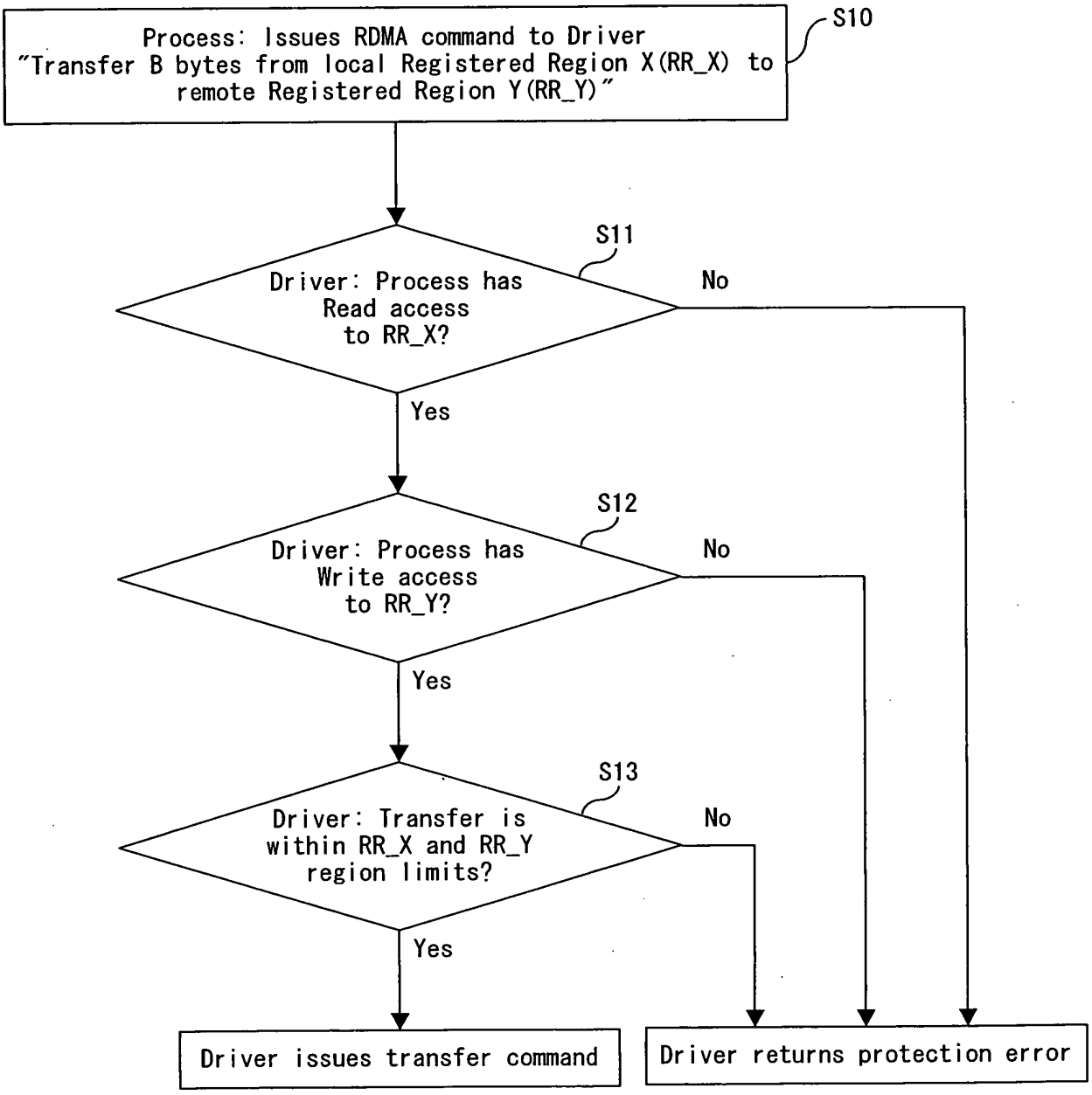
Process: Issues RDMA command to Driver "Transfer B bytes from local Registered Region X(RR_X) to remote Registered Region Y(RR_Y)"
S10
Driver: Process has Read access to RR_X?
S11
No
Yes
Driver: Process has Write access to RR_Y?
S12
No
Yes
Driver: Transfer is within RR_X and RR_Y region limits?
S13
No
Yes
Driver issues transfer command
Driver returns protection error

FIG. 6

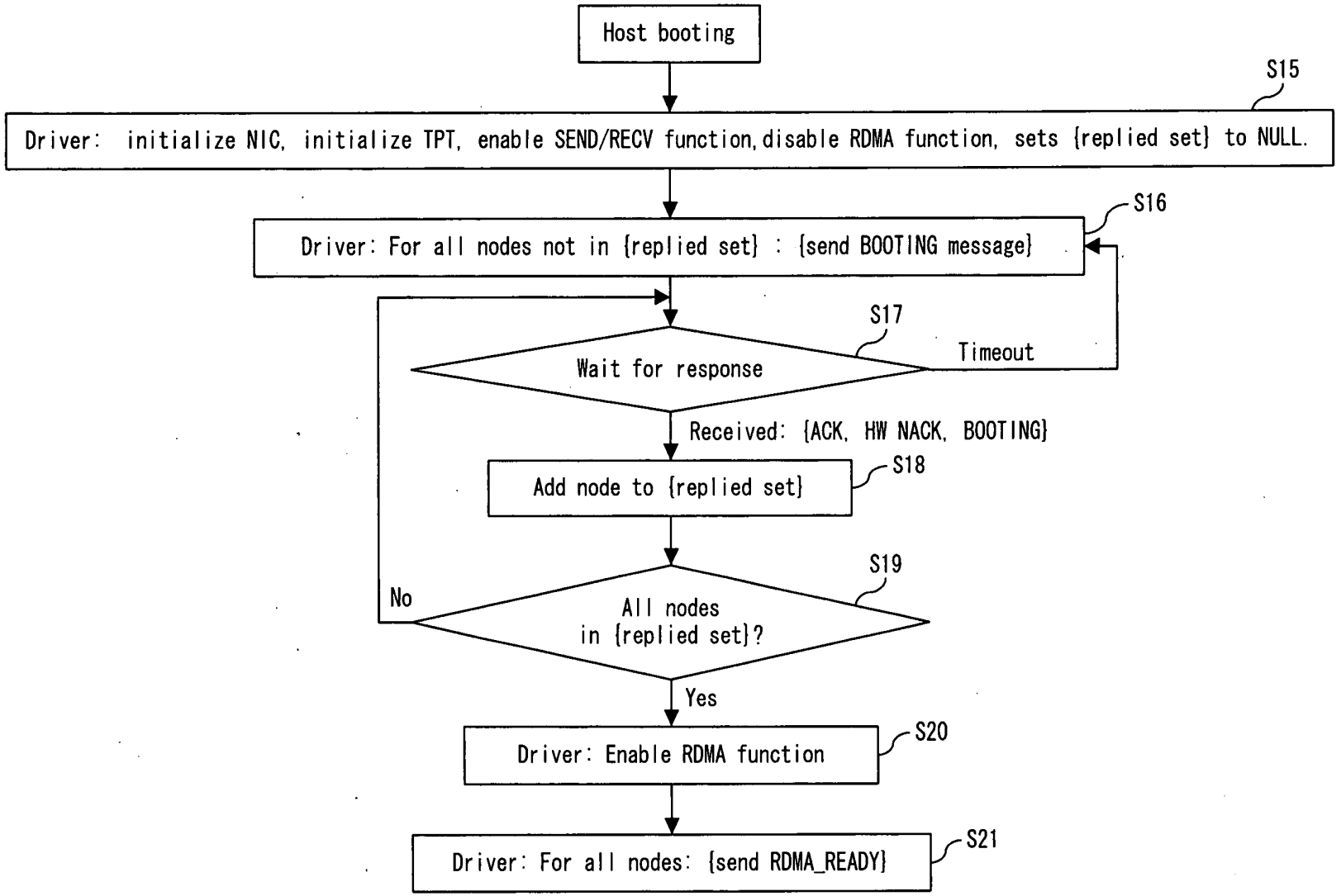
Host booting
S15
Driver: initialize NIC, initialize TPT, enable SEND/RECV function, disable RDMA function, sets {replied set} to NULL.
S16
Driver: For all nodes not in {replied set} : {send BOOTING message}
S17
Wait for response
Timeout
Received: {ACK, HW NACK, BOOTING}
S18
Add node to {replied set}
S19
No
All nodes in {replied set}?
Yes
S20
Driver: Enable RDMA function
S21
Driver: For all nodes: {send RDMA_READY}

FIG. 7

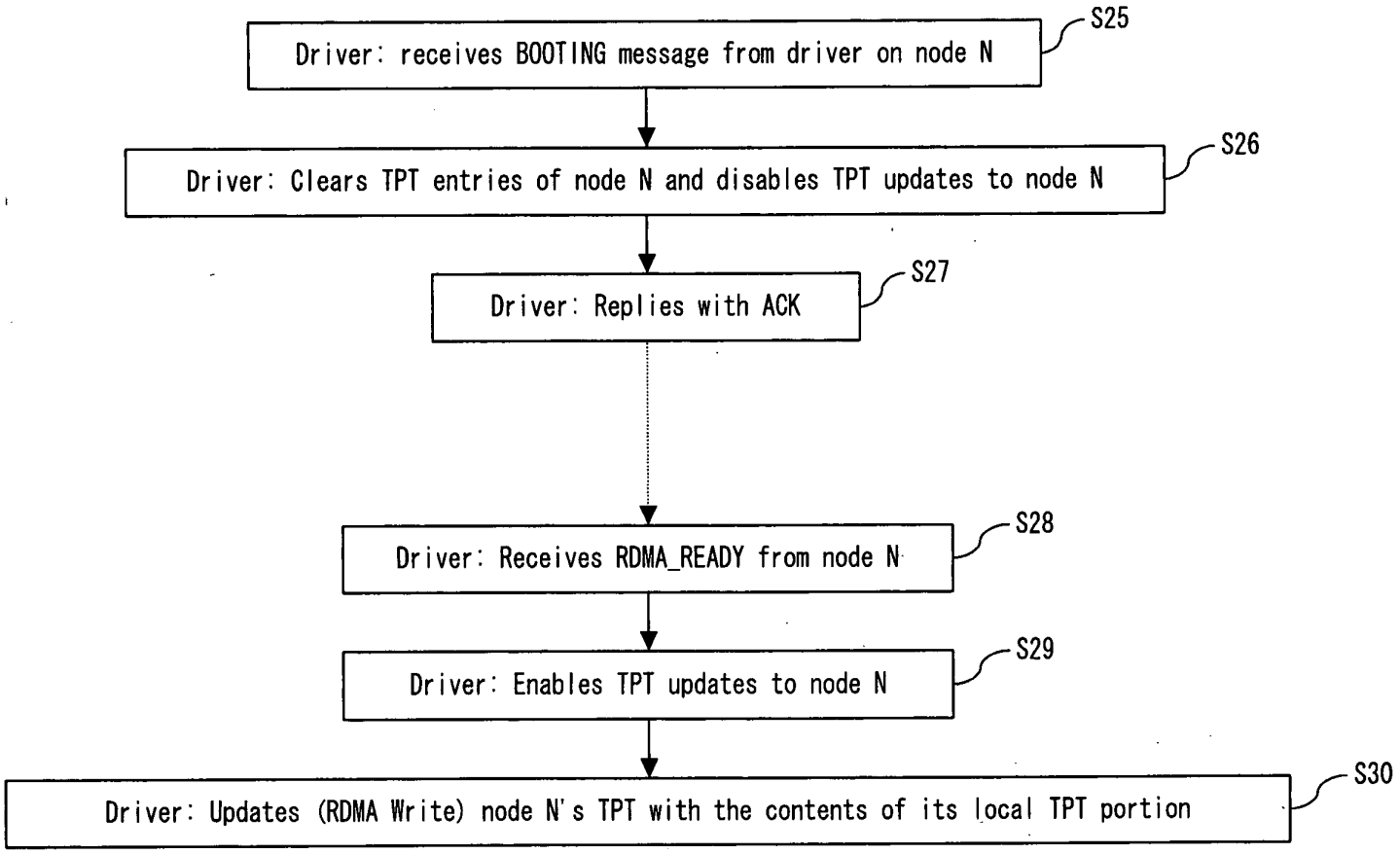
Driver: receives BOOTING message from driver on node N
S25
Driver: Clears TPT entries of node N and disables TPT updates to node N
S26
Driver: Replies with ACK
S27
Driver: Receives RDMA_READY from node N
S28
Driver: Enables TPT updates to node N
S29
Driver: Updates (RDMA Write) node N's TPT with the contents of its local TPT portion
S30

FIG. 8